ICS 03.060
CCS A11

团 体 技 术 报 告

TR/CBA 106—2022

人工智能模型风险管理框架

Framework for risk management of artificial intelligence model

2022－10－24 发布　　2022－10－24 实施

中国银行业协会　　发 布

目　　次

前　言

中国银行业协会(China Banking Association，CBA)成立于2000年5月，是经中国人民银行和民政部批准成立，并在民政部登记注册的全国性非营利社会团体，是中国银行业自律银行业金融机构。2003年中国银监会成立后，中国银行业协会主管单位由中国人民银行变更为中国银监会。2018年3月，中国银行保险监督管理委员会成立后，中国银行业协会主管单位由中国银监会变更为中国银行保险监督管理委员会。凡经业务主管单位批准设立的、具有独立法人资格的银行业金融机构(含在华外资银行业金融机构)和经相关监管机构批准、具有独立法人资格、在民政部门登记注册的各省(自治区、直辖市、计划单列市)银行业协会以及相关监管机构批准设立，具有独立法人资格的依法与银行业金融机构开展相关业务合作的其他类型金融机构，以及银行业专业服务机构均能申请加入中国银行业协会成为会员单位。

中国银行业协会日常办事机构为秘书处。秘书处设秘书长1名，副秘书长若干名。根据工作需要，中国银行业协会设立32个专业委员会，其中银行业产品和服务标准化专业委员会旨在开展银行业产品和服务标准化工作，包括制定和发布银行业的产品和服务标准，积极参与制定国家标准、行业规划，参与制定有关政策和法律法规，不断提高银行业产品和服务质量。

本文件按照T/CBA 1—2021《中国银行业协会团体标准化文件的结构和起草规则》的规定起草。

请注意本文件的某些内容可能涉及专利。本文件的发布机构不承担识别这些专利的责任。

本文件由中国工商银行股份有限公司、中国农业银行股份有限公司、交通银行股份有限公司、招商银行股份有限公司、广发银行股份有限公司、浙江网商银行股份有限公司联合提出。

本文件由中国银行业协会银行业产品和服务标准化专业委员会归口。

本文件起草单位：中国工商银行股份有限公司、中国农业银行股份有限公司、交通银行股份有限公司、招商银行股份有限公司、广发银行股份有限公司、浙江网商银行股份有限公司、上海浦东发展银行股份有限公司、兴业银行股份有限公司、渤海银行股份有限公司、中央国债登记结算有限责任公司、中国信息通信研究院。

本文件主要起草人：万芊、宋佳坤、魏凯、杨玲玲、陈杨、刘龙、牛菲菲、赵正龙、李曹建、李嘉铭、李金龙、文俊杰、赵文婷、蔡子龙、邓俊峰、谭蕾、彭晋、林冠辰、陆碧波、康平、何平、何海清、姜超、徐烨、胡铎镪、杨洋、董纪伟、李喆。

地 址：北京市西城区金融街 20 号交通银行大厦 B 座 11-12 层
电 话：010-66553368 010-66291132
邮 编：100033
邮 箱：cba.china@china-cba.net
传 真：010-66553356

引　言

人工智能技术在银行业的应用是银行业金融技术数字化转型的重要手段，在产生效益的同时，也带来了新的风险和挑战。如模型数据来源、准确性和及时度不当产生的数据风险，模型缺陷造成的决策失衡和结果偏差，模型安全和监控缺失造成的风险事件等。

本文件针对银行业金融机构在管理、研发、供应、使用人工智能模型过程中面临的主要风险挑战，根据行业特性，给出银行业金融机构在开展人工智能模型需求分析、数据准备、模型构建、检验验证、模型部署、模型监控、持续验证与确认、模型修正、模型下线等关键活动过程中进行风险管理需要考虑内容的框架。

由于业务目标的不同、应用场景的不用、模型使用技术、银行业金融机构风险偏好的不同等因素，不同的人工智能模型所面临的威胁有所不同，风险管理需求也会有所差异。为了实现对不同风险管理需求的人工智能模型的管理，本文件中的人工智能模型风险管理能力分为三个层级。若本文件描述的各个方面能按照银行内部的管理体系和管理惯例反映在其内控控制的相关文件中并提出判定和控制措施，将有助于本文件的实施者快速灵活应对模型风险，更好地将人工智能技术应用于数字化转型。

人工智能模型风险管理框架

1 范围

本文件提供了银行业金融机构在开展人工智能模型需求分析、数据准备、模型构建、检验验证、模型部署、模型监控、持续验证与确认、模型修正、模型下线等关键活动过程中进行风险管理考虑的相关方面。

本文件适用于管理、研发、供应、使用人工智能模型的银行业金融机构。向银行业金融机构提供人工智能服务和支持的相关机构亦能作为参考。

2 规范性引用文件

下列文件中的内容通过文中的规范性引用而构成本文件必不可少的条款。其中，注日期的引用文件，仅该日期对应的版本适用于本文件；不注日期的引用文件，其最新版本（包括所有的修改单）适用于本文件。

JR/T 0101—2013 银行业软件测试文档规范

3 术语和定义

下列术语和定义适用于本文件。

3.1

模型 model

系统、实体、现象、过程或数据的物理、数学或其他逻辑表示

[来源：ISO/IEC TR 24030:2021，3.31]

3.2

人工智能模型 artificial intelligence model

使用一种或多种人工智能技术和方法构建的模型（3.1）。

术语条目注 1：人工智能模型能够针对一组给定的人类定义的目标生成输出，例如内容、预测、建议或影响他们互动环境的决策。

3.3

风险 risk

不确定性对目标的影响

[来源：ISO 31000:2018，3.1]

3.4

模型风险 model risk

因模型（3.1）自身缺陷形成不准确的输出、错误所导致的风险（3.3）或不恰当地使用模型所导致的风险。

4 缩略语

下列缩略语适用于本文件。

AI	人工智能（Artificial Intelligence）
PSI	群体稳定性指标（Population Stability Index）
IV	信息价值（Information Value）
WOE	证据权重（Weight of Evidence）
KS	柯斯统计量（Kolmogorov-Smirnov）
ROC	接受者操作特征（Receiver Operating Characteristic）
AUC	ROC曲线下的面积（Area Under Curve）
GINI	基尼系数（Gini Coefficient）

5 风险管理等级划分

5.1 基本策略

由于银行业金融机构应用人工智能模型的业务领域、重要性和影响程度、风险敞口不同，且银行业金融机构的规模和风险偏好也存在差异，故各银行业金融机构针对人工智能模型的风险管理需求也不相同。为了适用于对人工智能模型具有不同风险管理需求的银行业金融机构，本文件按“分级管理、逐步递增”的策略进行组织。

5.2 基础等级

本文件中的人工智能模型风险管理措施分为逐步递增的三个级别。

a)第一级，银行业金融机构未在任何业务建立模型分级方法/流程，仅根据临时需求或基于个人经验，对模型进行分级。银行业金融机构在管理、研发、供应、使用人工智能模型的过程中不能有效地控制风险，仅在部分过程中根据临时的需求执行了相关工作，或相关人员基于经验开展活动。

b)第二级，银行业金融机构从业务和技术层面，由不同内设单元对管理、研发、供应、使用人工智能模型过程中主要的、常见的风险进行短期计划和现场控制；建立根据模型影响的对象和影响程度，对业务范围内的模型进行分级管理；分配了风险管理资源，明确了必要的责任；模型进行分级标识和管理；现有的风险控制措施执行情况有记录能查。

c)第三级，银行业金融机构明确模型分级原则、方法和操作要求，在企业级对人工智能模型风险管理进行全面规划、计划、管控、审计，建立了覆盖全业务领域、全工作流程、全参与角色的企业标准体系；对不同级别的模型建立相应的安全管理要求和管理措施；对标准体系实施有全面和完整的记录，对执行的效果有评估和审计，对发现的问题进行了有效的问题分析，并与绩效考核挂钩。

5.3 等级细化与拓展

银行业金融机构根据管理需要和技术发展趋势，对5.2中的人工智能模型风险管理等级能够进行进一步的细化与拓展。

所有的细化内容不能超出本级的要求，细化后再行概括，能够还原为原等级要求；所有的拓展内容高于本级原有要求，但不能与现有要求存在不一致。

6 模型生命周期

6.1 需求分析

6.1.1 第一级

银行业金融机构未对模型需求进行有效管理，团队相关人员基于经验开展基本的需求分析活动。

6.1.2 第二级

第二级需求分析工作如下。

a) 银行业金融机构内设部门建立需求管理流程，规范需求编写、需求变更、需求评审、需求验收等环节。

b) 遵循需求管理流程，明确以下内容：
   1) 模型需求提出、承接、评审、确认的角色并落实到具体部门或人员；
   2) 模型需求提出、承接、评审、确认的具体流程、操作规范和要求；
   3) 模型需求过程中的关键文档，包含但不限于需求说明书；
   4) 模型需求的版本管理方式，包含但不限于需求基线和需求变更。

c) 编写模型需求说明文档，包含以下内容：
   1) 模型的业务需求背景、需求调研和需求描述；
   2) 模型应用的目标和方式，包含模型效果以及效果收集形式、模型评价方式及评价指标、应用场景、局限性、合理性；
   3) 模型的业务价值和技术价值；
   4) 模型影响，包含模型调用量级、影响用户类型和规模。

d) 参与需求评审的相关方至少包括模型使用方、模型开发方。

e) 形成模型需求评审记录，包含评审内容、评审结果和相关方确认。

6.1.3 第三级

第三级需求分析工作如下。

a) 6.1.2 的全部内容。

b) 设立需求管理人员，负责模型需求的承接、分析、排序、分发。

c) 需求管理流程遵循银行业金融机构治理与管理要求。

注：在大部分银行业金融机构内，银行业金融机构的治理与管理要求通过组织架构与规章制度体现。

d) 模型需求评审的内容包含业务需求分析和拆解的准确性、模型方案的可行性、模型的安全、合规、风险以及缓释措施的有效性。

e) 明确参与模型需求评审的相关方及各方在评审过程中评审的重点内容，相关方具备专业知识并能够基于具体的模型进行评估和审核。

f) 需求评审的参与方包含独立于模型开发方和模型使用方的部门人员，各参与方向不同的管理层人员进行整体情况汇报。

g) 部署需求管理平台或工具，支持模型需求工作流管理(包括需求变更、排序等)、提供模型需求文档流转和查阅、支持需求具体部门和人员角色设置、保留需求阶段重要操作日志和记录。

h) 模型上线版本与需求保持一致。

6.2 数据准备

6.2.1 第一级

银行业金融机构未对模型开发所使用的数据进行有效管理，模型开发相关人员基于经验开展基本的数据准备活动。

6.2.2 第二级

第二级数据准备工作如下。

a) 银行业金融机构内设部门建立模型数据准备管理流程，明确以下内容：
   1) 开展数据准备的角色并落实到具体下属部门或人员。
   2) 数据质量检查准则、检查方法、检查维度、样本量和样本范围、样本保留时间和方式。
   3) 数据准备过程中的关键文档，包含但不限于数据质量检查结果记录。
b) 记录数据准备过程，包括数据获取时间、训练数据来源、训练数据量、数据存储介质标识、采样方法。
c) 用于模型开发和测试的样本保留周期不少于 6 个月。
d) 模型开发所使用的数据在建模前（生产数据接入特征平台前）经过模型使用方负责人、模型开发方、数据所有者、数据安全管理人员等审核。
e) 通过权限申请和审核批准，对模型开发所使用的数据的访问和使用进行权限管控，保证数据的访问和使用限于该模型开发所需的最小范围。
f) 数据准备环境中如包含敏感个人信息，采取屏蔽等技术措施防止数据展示可能导致的信息泄露。
g) 部署数据开发平台或工具，支持以下功能：
   1) 数据集成、数据清洗和数据加工等操作；
   2) 记录数据准备过程；
   3) 记录人员操作；
   4) 记录数据开发日志；
   5) 记录数据开发环境；
   6) 记录变量结构和变量处理过程；
   7) 支持数据申请和审批。
h) 保留和备份数据开发日志信息，保存时长不少于 6 个月。

6.2.3 第三级

第三级数据准备工作如下。

a) 6.2.2 的全部内容。
b) 数据准备流程遵循银行业金融机构治理与管理要求。
c) 针对模型的预期用途进行专有的数据开发，考虑特定的应用场景所特有的特征或要素。
d) 对用于模型开发和测试的数据的质量进行检查，包含以下内容：
   1) 对数据的准确性、完整性、一致性和全面性进行检查；
   2) 通过勾稽关系检查、横向比较、趋势分析进行检查；
   3) 对缺失值、异常值、极端值等进行检查。
e) 当扩展新的数据维度、显著扩大数据量级或模型应用于不同于原定业务场景时，重新进行模型构建。
f) 利用角色、流程、权限设置对训练数据集和测试数据集进行管控，保证数据集的完整性和一致性。
g) 利用角色、流程、权限设置对特征进行管控，保证特征的完整性和一致性。
h) 数据准备环境中如包含敏感个人信息，采取加密等技术措施保证数据存储的保密性。
i) 数据开发平台或工具在第二级的基础上，支持以下功能：
   1) 记录数据训练历史脚本、数据样本、训练时间等；
   2) 支持模型历史训练数据的记录能追溯。

6.3 模型构建

6.3.1 第一级

银行业金融机构未对模型构建过程进行有效管理，模型开发人员基于经验开展基本的模型构建活动。

6.3.2 第二级

第二级模型构建工作如下。

a) 银行业金融机构内设部门建立模型构建管理流程，落实以下内容：
   1) 开展模型构建的角色并落实到具体下属部门或人员（数据开发者）；
   2) 模型构建的具体流程、要求和操作规范，包括模型构建过程中遵循的安全原则和安全要求，模型方法探索过程和比较方式、模型性能比较方式；
   3) 模型评审的具体流程、要求和操作规范；
   4) 模型构建过程中的关键文档，包含但不限于模型开发文档。
b) 编写模型开发文档，包含以下内容：
   1) 模型背景；
   2) 模型用途和目标；
   3) 模型相关方，包括模型开发者、模型使用方、模型所有方、模型运维者；
   4) 模型相关假设（如有）的说明和测试记录；
   5) 模型的弱点和局限性；
   6) 模型方法探索过程、比较方式和历史结果；
   7) 模型性能比较方式和历史结果；
   8) 样本定义和变量设计；
   9) 数据描述和分析；
   10) 模型需求设计方案；
   11) 建模变量清单；
   12) 模型结果（参数选择）；
   13) 模型评估；
   14) 数据样本和样本生成；
   15) 模型监控方案；
   16) 资源投入和时间安排。
c) 评估模型应用可能产生的合规风险、声誉风险、业务风险、财务风险、用户权益危害等，调研和分析能采取的风险控制措施，将模型应用风险降低至能接受水平。
d) 对于可解释模型，将模型的输入特征与模型的输出的结果建立联系。能实现对于给定的模型输出，能关联到与该输出最为相关的输入特征中。

注：银行业金融机构所使用的模型必须符合《中国银保监会办公厅关于银行业保险业数字化转型的指导意见》的要求。

e) 银行业金融机构包括模型开发者、模型使用方、数据所有者（如果数据直接取自银行业金融机构内部已加工整合的数据资产，可不需数据所有者参与）等在内的人员对模型进行评审。
f) 形成模型评审记录，至少包含评审内容、评审结果和相关方确认。
g) 部署模型构建平台或工具，支持以下功能：
   1) 样本管理；
   2) 特征管理；
   3) 记录单个特征的详细信息；
   4) 模型评估；

5） 记录模型构建过程；
6） 记录人员操作；
7） 记录模型构建日志；
8） 记录模型构建环境；
9） 支持精细化的权限申请和审批。

h） 保留模型开发代码。
i） 保留和备份模型构建日志信息，保存时长不少于 6 个月。

6.3.3 第三级

第三级模型构建工作如下。

a） 6.3.2 的全部内容。
b） 模型构建流程遵循银行业金融机构治理与管理要求。
c） 如进行模型修正，修订模型开发文档、更新相关内容，并说明新模型相较于旧模型的优越性。
d） 编写代码编写规范和安全规范，涵盖模型代码开发常见的危险函数、算法框架、安全漏洞等。
e） 使用代码扫描工具对代码进行评审，包括但不限于合规和安全缺陷检查，如发现安全缺陷问题，以工具直接修复或工具结合人工的方式修复。
f） 模型构建平台或工具在第二级的基础上，支持以下功能：
1） 记录模型方法探索过程、比较方式和历史结果；
2） 记录模型性能比较方式和历史结果；
3） 具备模型文件校验等技术能力，支持对模型文件格式、大小、参数范围、节点名称、数据维度等关键信息进行检测校验；
4） 适用于多种用户角色的建模工具，如给业务人员提供图形化的建模工具、给模型开发方提供专业的交互式建模工具等；
5） 提供的算法库能够支持多种模型的构建需求，包括有监督、半监督、无监督，以及主流的神经网络、深度学习等。
g） 对于可解释模型，在建立模型输出与输入特征的关联基础上，解释结果能捕捉到特征与特征之间的联系，并展示有关联的特征如何共同作用于模型输出结果。
h） 根据该应用场景构造特征。
i） 根据该应用场景构建模型。
j） 成立模型评审委员会或独立于模型开发方的模型专业人员对模型进行评审。
k） 对模型使用过程中产生的相关计算数据进行保护，包括输出向量、模型参数、模型梯度等可能会泄露训练数据的敏感信息或者模型自身的属性参数。如限制恶意访问次数、引入随机性、添加模型水印等。
l） 对目标函数进行说明，目标函数设计上不存在针对特殊群体的偏见歧视。
m） 制定适用于不同类型和级别的模型安全和风险评估清单，并根据清单开展模型安全和风险影响评估。评估清单包含以下内容：
1） 数据合规风险；
2） 数据质量风险；
3） 目标函数安全；
4） 模型（算法）选择风险；
5） 算法依赖库风险；
6） 算法使用风险；
7） 模型安全漏洞。

6.4 模型验证

6.4.1 第一级

银行业金融机构未对模型验证进行有效管理，模型验证相关人员基于经验开展基本的测试和验证活动。

6.4.2 第二级

第二级模型验证工作如下。

a) 银行业金融机构内设部门建立模型验证管理流程，规范以下内容：
   1) 开展模型验证的角色并落实到具体部门或人员；
   2) 执行模型验证的时机；
   3) 模型验证的范围和方法；
   4) 模型验证发现的问题分类分级定义；
   5) 问题处置、追踪和汇报的方式；
   6) 问题解决的定义；
   7) 模型验证的结论定义（如适合使用、有条件使用、不适合使用）；
   8) 模型验证过程中的关键文档，包含但不限于模型验证报告。
b) 模型验证包含以下内容：
   1) 模型需求设计，即模型方法的合理性、模型局限性和弱点、模型预期的用途等；
   2) 模型开发过程，即需求分析、数据准备、模型构建等过程准确合规可控；
   3) 模型效果，即对模型输出与相应的实际结果进行比较，包括评价估算或预测的准确性、评价排序能力等。
c) 模型部署前进行集成测试、配置项测试、系统测试、验收测试，编写并保留测试文档。测试文档规范按 JR/T 0101-2013 中的要求实施。
d) 编写模型验证报告，包含以下内容：
   1) 模型基本信息，包括模型名称、模型版本、模型背景和用途、模型方法等；
   2) 模型相关方，包括模型开发者、模型使用方、模型所有方、模型验证者；
   3) 模型验证时间；
   4) 模型验证方式；
   5) 模型验证方案；
   6) 问题发现和影响评估；
   7) 模型验证结论。
e) 如进行模型修正，在部署之前重新验证。

6.4.3 第三级

第三级模型验证工作如下。

a) 6.4.2 的全部内容。
b) 设置模型验证专职人员，与模型使用方和模型开发方向不同的高级管理层人员汇报。
c) 模型验证流程遵循银行业金融机构治理与管理要求。
d) 遵循模型验证管理流程，制定适用于不同类型和级别的模型验证准则和检查矩阵，规范模型验证的具体内容。包含验证类别、验证项、通过条件、风险等级等。要求模型验证人员参照执行。
e) 同时采用自动化验证和人工验证方式进行交叉验证。
f) 模型验证在第二级的基础上，包含以下内容：

1） 模型的健壮性验证，也可称为鲁棒性验证，重点验证模型对数据变化的容忍度，支持模型入参的敏感性验证，检测输入和参数值的微小变化对模型输出的影响；
2） 建模数据的稳定性和真实性，以及数据的适当性、合理性、权重比例的科学性；
3） 数据安全合规，即数据处理过程中敏感数据是否通过加密、去标识化、匿名化等方式进行保护，数据处理结果中是否存在可恢复或者涉及安全隐私的敏感数据；
4） 一定条件下的公平性验证，即模型决策结果保证人群均等、机会均等、几率均等、人群无关性等，包括显性偏见和隐形偏见识别；
5） 采用模拟数据窃取、成员推理攻击、数据逆向还原等方法进行验证；
6） 可复现性验证，即在相同场景下，采取不同的数据集来对模型进行多次校验，并对在相同条件下出现的差异化结果进行分析；
7） 通过同时改变多个输入来发现意外的交互作用，特别是在交互作用复杂且不直观的情况下。

g） 部署模型验证平台或工具，支持以下功能：
1） 数据质量验证；
2） 模型自动化验证；
3） 记录验证人员操作；
4） 记录模型验证环境；
5） 记录模型验证历史脚本、数据样本、验证时间、验证结果。

6.5 模型部署

6.5.1 第一级

银行业金融机构未对模型部署进行有效管理，模型部署相关人员基于经验开展基本的部署活动。

6.5.2 第二级

第二级模型部署工作如下。

a） 银行业金融机构内设部门建立模型部署管理流程，明确以下内容：
1） 执行模型部署的角色并落实到下属部门或人员；
2） 模型上线的具体流程与要求，包括模型上线前评审、模型校验、生产数据接入；
3） 模型上线前评审的关键文档。

b） 评估模型部署影响范围，提前将模型部署可能造成的影响告知相关方。

c） 记录模型部署过程，包括模型部署的操作人员、部署环境、部署步骤、部署时间和部署结果。

d） 执行模型上线前评审，形成模型评审记录，至少包含评审结果和相关方确认。

e） 模型上线前评审包含以下内容：
1） 模型业务应用；
2） 模型技术方案；
3） 模型开发过程；
4） 模型部署回退方案；
5） 模型调用方案。

f） 模型工程化部署，包含以下内容：
1） 数据源调整；
2） 参数化处理；
3） 监控指标设定；
4） 运行效率评估优化；

5） 多模型组合服务。

g） 部署模型部署平台或工具，支持以下功能：

1） 记录部署人员操作；

2） 记录部署时间和部署结果；

3） 记录部署相关脚本；

4） 记录部署软硬件环境和配置信息；

5） 执行模型文件完整性校验；

6） 支持暂停上线和上线终止；

7） 支持安全检测和访问控制。

h） 保留和备份模型部署日志信息，保存时长不少于 6 个月。

6.5.3 第三级

第三级模型部署工作如下。

a） 6.5.2 中的全部内容。

b） 模型部署流程遵循银行业金融机构治理与管理要求。

c） 通过模型灰度发布对模型进行试运行，确保模型效果符合要求后发布模型。

d） 模型上线前评审由模型开发方、模型使用方和独立于前两者的第三方参与，任一评审参与方均能否决模型上线。

e） 制定模型上线前评审表，规范模型上线前评审各参与方评估的主要内容和具体评审要求。

f） 模型部署平台或工具在第二级的基础上，支持以下功能：

1） 自动识别模型部署影响范围并告知相关方；

2） 支持模型灰度发布；

3） 支持模型流量分配（支持异步全量和同步分流）；

4） 实时监控模型部署过程产生的风险并能够可视化展示；

5） 支持回滚至前一版本；

6） 支持自动化、精细化、多样化的模型部署方式；

7） 支持下载管理，避免将模型文件直接下载到本地。

6.6 模型监控

6.6.1 第一级

银行业金融机构未对已上线模型进行有效管理，相关人员基于经验开展基本的运行监控。

6.6.2 第二级

第二级模型监控工作如下。

a） 银行业金融机构内设部门建立模型监控管理流程，规范以下内容：

1） 执行模型监控的角色并落实到具体下属部门或人员；

2） 模型监控策略、监控内容、监控方式和评价方法、风险预警和响应措施等；

3） 模型监控过程中涉及的关键文档。

b） 模型使用方、模型所有方、模型运维者等相关方，基于模型应用的业务目标和风险，根据各自职责设置模型运行监控指标。

c） 模型所有方记录模型的使用者和使用场景，并留存记录。

d） 模型监控包含以下内容：

1） 数据监控，即监控模型所需数据的提取、加工和数据分布的变化；
2） 模型运行监控，模型运行正常与否、模型运行的资源消耗、模型运行响应的性能（如单笔响应时长、总调度量）等；
3） 结果监控，即监控模型直接输出数据的变化。

e） 定期识别和量化模型效果或风险，对由模型效果降低带来的风险进行预警。
f） 定期开展监控策略的合理性评估，及时调整运行监控策略。
g） 制定模型性能监测计划，包括监测频率、指标、基准点和指标的阈值。
h） 定期产出监控报告，包含以下内容：
1） 模型性能表现；
2） 模型支持体系的运行情况；
3） 模型运行环境或假设条件的变化对模型结果的影响。
i） 部署模型监控平台或工具，支持以下功能：
1） 基础监控指标配置；
2） 触发预置告警；
3） 输出模型监测评价报告；
4） 支持模型下线；
5） 支持可视化的风险展示；
6） 多种方式的预警通知。
j） 保留和备份模型监控日志信息，保存时长不少于 6 个月。

6.6.3 第三级

第三级模型监控工作如下。

a） 6.6.2 中的全部内容。
b） 模型监控流程遵循银行业金融机构治理与管理要求。
c） 设置团队/人员对模型的监控进行独立审查，该团队/人员与模型使用方和模型开发方向不同的高级管理层汇报。
d） 根据模型决策重要性和风险影响制定差异化的监控和响应策略。高风险模型产生的结果应用时需考虑人工检查、人机交叉验证等处理策略。中风险模型决策结果突破风险阈值后由人工进行处置。
e） 模型监控平台或工具在第二级的基础上，支持以下功能：
1） 灵活配置监控指标，支持不同模型类型的不同性能监控，建立与模型类别、使用用途相适配的模型监控指标，例如：PSI、IV、WOE、KS、ROC、AUC、GINI、卡方检验、F 检验、T 检验、秩和检验等常见模型指标监控；

注：PSI是模型稳定性指标，反映了验证样本在各分数段的分布与建模样本分布的稳定性。

IV是特征预测能力指标，一般用来表示特征对目标预测的贡献程度，即特征的预测能力。

WOE是特征预测能力指标，指对于字符型变量的某个值或者是连续变量的某个分段下的正负样本的比例的对数。

KS是模型区分度指标，表示正负样本累计分布之间的差值。

GINI是模型区分度指标，GINI系数越大，表明模型对正负样本的评估差异性越大，模型的区分能力越强。

ROC曲线指接受者操作特征曲线，是根据一系列不同的二分类方式，以真阳性率为纵坐标，假阳性率为横坐标绘制的曲线。

AUC指ROC曲线下的面积，用来衡量模型的预测效果。

卡方检验表示统计样本的实际观测值与理论推断值之间的偏离程度。

F检验指方差比率检验、方差齐性检验，是一种在零假设之下，统计值服从F-分布的检验。

T检验指用T分布理论来推论差异发生的概率，比较两个平均数的差异是否显著。

秩和检验又称顺序和检验，是一种非参数检验，不依赖于总体分布的具体形式，应用时可以不考虑被研究对象为何种分布以及分布是否已知。

2） 对模型相关方反馈的持续监测；

3） 监控的应急处理；

4） 灵活的处置策略，如熔断、降级、隔离、标记、模型（自动）更新等；

5） 人工进行部分或全部干预和处置；

6） 对模型业务成效进行监控。基于业务场景采用不同的业务成效监控指标进行监测和评估。

6.7 持续验证与确认

6.7.1 第一级

银行业金融机构未对模型持续验证与确认进行有效管理，模型验证相关人员基于经验开展持续验证与确认活动。

6.7.2 第二级

第二级持续验证与确认工作如下。

a） 银行业金融机构内设机构建立模型持续验证与确认管理流程，规范以下内容：

1） 开展模型持续验证与确认的角色并落实到具体部门或人员；

2） 模型持续验证与确认频次或周期，确保能够应对及时性风险；

3） 持续验证与确认的范围和方法等；

4） 持续验证与确认过程中的关键文档。

b） 由模型的所有者或使用者开展持续验证与确认。

c） 持续验证与确认过程的各项文档输出与模型验证过程一致。

d） 开展持续模型验证考虑以下内容：

1） 结合模型运行监控内容；

2） 模型的方法、假设、局限性和弱点等；

3） 模型用途和使用人群是否发生变化；

4） 监管合规和市场状况是否发生变化；

5） 使用最新生成的数据进行持续验证与确认。

e） 持续验证与确认包含以下内容：

1） 模型验证环节执行的验证内容；

2） 模型监控指标的完备性、合理性和有效性；

3） 发生的模型重大技术风险事件和业务风险事件。

6.7.3 第三级

第三级模型持续验证与确认工作如下。

a） 6.7.2 中的全部内容。

b） 持续验证与确认流程遵循银行业金融机构治理与管理要求。

c） 模型验证专职人员开展持续确认。

d） 结合外部舆情监测内容进行模型相关的持续验证与确认。

e） 同时采用自动化验证和人工验证方式进行交叉验证。

f） 遵循模型验证管理流程，制定适用于不同类型和级别的模型验证准则和检查矩阵，规范模型验证的具体内容。包含验证类别、验证项、通过条件、风险等级等。要求模型验证人员参照执行。

g） 具备在极大偶然性、随机性、无模型历史可循且具有大量不确定性的决策不能重复情况下，识别和处理模型异常的方法。

6.8 模型修正

6.8.1 第一级

银行业金融机构未对模型修正进行有效管理，模型修正相关人员基于经验开展模型修正活动。

6.8.2 第二级

第二级模型修正工作如下。

a） 银行业金融机构内设机构建立模型修正管理流程，规范以下内容：

1） 开展模型修正的角色并落实到具体下属部门或人员；

2） 模型修正触发条件，即需要进行模型修正的情况；

3） 模型修正的具体流程、要求和操作规范；

4） 规范模型修正周期、修正范围、修正指标等；

5） 模型修正过程中的关键文档，包含但不限于模型修正方案、模型修正记录、模型开发文档。

b） 发生以下情况时，触发模型修正：

1） 数据源发生变化；

2） 模型性能指标突破阈值；

3） 模型越控且长期需要人工修正；

注：越控是指在使用任何模型时，都会出现模型输出结果基于模型使用者的专家判断被忽略、更改或被反转的情况。越控一定程度表明了模型在某些方面没有按预期表现或存在局限的情况。

4） 监管合规和业务需求发生变化。

c） 模型修正前编写模型修正方案，包含以下内容：

1） 模型基本信息，包括模型名称、模型版本、模型背景和用途、模型方法等；

2） 模型相关方，包括模型开发者、模型使用方、模型所有方、模型验证者；

3） 模型修正的背景和原因；

4） 模型修正使用的数据；

5） 模型修正范围；

6） 执行模型修正的角色和职责；

7） 模型修正实施步骤。

d） 保留模型修正记录，包含以下内容：

1） 模型修正前后版本；

2） 模型修正前后参数；

3） 模型修正前后样本类型、时间窗口等；

4） 模型修正前后性能数据。

e） 部署模型修正平台或工具，支持以下功能：

1） 模型版本管理；

2） 记录模型修正过程；

3） 记录人员操作；

4） 记录模型修正日志。

f） 对模型越控及进行记录。分析模型越控原因，跟踪和评估越控效果，采取适当的应对措施。

6.8.3 第三级

第三级模型修正工作如下。

a） 6.8.2 中的全部内容。
b） 模型修正流程遵循银行业金融机构治理与管理要求。
c） 根据模型修正情况更新模型开发文档，涵盖模型修正前后的各项参数对比记录。
d） 模型修正完成后与原有模型进行 A/B 测试验证。
e） 模型修正后通过模型验证、部署环节重新上线。
f） 部署模型修正平台或工具在满足第二级的基础上，支持模型自动检查和自动监控。

6.9 模型下线

6.9.1 第一级

银行业金融机构未对模型下线进行有效管理，模型下线相关人员基于经验开展模型下线活动。

6.9.2 第二级

第二级模型下线工作如下。

a） 银行业金融机构内设机构建立模型下线管理流程，明确以下内容：
   1） 开展模型下线的角色并落实到下属部门或人员；
   2） 进行模型归档的角色并落实到下属部门或人员；
   3） 明确模型下线条件、模型下线需求确认、模型下线影响评估、模型下线审核、模型回滚和下线后评估等具体流程和要求；
   4） 模型归档的具体流程和要求；
   5） 模型下线、归档过程中的关键文档。
b） 由系统运维人员执行下线操作。
c） 评估模型下线影响范围，确认无下游调用，制定应急方案，并提前将模型下线可能造成的影响告知相关方。
d） 记录模型下线过程，包括模型下线的操作人员、下线步骤、下线时间和下线结果等。
e） 执行模型下线审核，与模型使用方确认下线需求，模型开发方（模型开发者）负责人审核下线。
f） 形成模型下线审核记录，至少包含审核结果和相关方确认。
g） 部署模型下线和归档平台或工具，支持以下功能：
   1） 支持多版本同时在线；
   2） 记录下线人员操作；
   3） 记录下线时间和下线结果；
   4） 支持下线审核；
   5） 记录下线模型软硬件环境和配置信息；
   6） 支持保留和查阅模型生命周期中所有关键的文档和记录。
h） 模型归档后建立模型档案，包括模型生命周期各阶段第二级要求形成的各类代码、数据、文档和记录。
i） 模型归档活动于模型下线后 3 个月内完成。

6.9.3 第三级

第三级模型下线工作如下。

a) 6.9.2 中的全部内容。
b) 模型修正流程遵循银行业金融机构治理与管理要求。
c) 模型下线后进行验证，确保下线模型与验证结果一致，并持续监测和评估对业务的影响。
d) 定期排查、下线不符合业务预期的、无效的或存在重大问题的模型，并形成排查记录。
e) 对下线的模型进行归档，制定对已归档模型及相关文档的管理要求。
f) 模型归档后建立模型档案，包括模型生命周期各阶段要求形成的全部代码、数据、文档和记录。
g) 模型下线平台或工具在第二级的基础上，支持以下功能：
   1) 检查模型结果的引用关系（模型血缘）；
   2) 自动化的方式管理模型下线；
   3) 支持直接下线、灰度发布下线、流量分配、回滚至上一版本等。
h) 根据模型下线的影响程度采用适当的下线方式，对业务影响较大的模型通过模型灰度发布下线。

7 特殊过程

7.1 模型外部合作

7.1.1 第一级

银行业金融机构未对模型外部合作进行有效管理，团队相关人员基于经验对外部合作进行管理。

7.1.2 第二级

第二级模型外部合作工作如下。

a) 合作开展前，对外部合作方开展尽职调查。
b) 模型使用方、模型开发者等相关方在开展外部合作前，对模型合作内容、合作模式、影响范围和影响程度，以及安全风险和控制措施的有效性、合法合规情况等进行评估，出具评估结论。
c) 定期对模型相关的外部合作进行检查。
d) 针对涉及高风险模型的外部合作制定应急预案，并定期进行演练。

7.1.3 第三级

第三级模型外部合作工作如下。

a) 7.1.2 中的全部内容。
b) 在模型部署前对涉及外部合作的内容进行安全风险评估，或提供由独立的第三方机构出具的评估报告。
c) 根据涉及外部合作模式不同，对模型进行评估，识别存在的缺陷和问题，形成评估报告。
d) 与外部合作方通过合同等形式明确双方的安全措施，包括在发生风险事件、合作中止或提前结束、用户权益纠纷等情况时双方的沟通渠道和响应机制。
e) 定期对外部合作相关的模型应急预案组织双方联动演练，出具演练报告。
f) 对非受控环境下的外部合作方提供的模型进行实时监控。

7.2 模型应急响应

7.2.1 第一级

银行业金融机构未对模型应急响应进行有效管理，团队相关人员基于经验开展基本的应急响应活动。

7.2.2 第二级

第二级模型应急响应工作如下。

a) 银行业金融机构内设部门制定业务应急响应方案，涵盖模型相关的应急事件类型、风险场景和应急处置措施。
b) 对可能对银行业金融机构或用户存在重大影响的处置动作设置决策机制。
c) 业务应急响应时限要求内执行人工或自动化模型下线操作。
d) 定期开展应急响应培训和演练。
e) 部署应急响应平台或工具，支持以下功能：
   1) 模型应急监控；
   2) 模型预警；
   3) 应急处置。

7.2.3 第三级

第三级模型应急响应工作如下。

a) 7.2.2 中的全部内容。
b) 建立模型应急事件分级规则，明确应急响应时效、同步时效、止血时效，设置决策机制。
c) 建立和维护模型应急预案，涵盖以下内容：
   1) 模型应急组织、人员及相关职责，包含系统运维人员、模型开发方人员、数据质量人员以及技术支持人员模型开发方；
   2) 模型风险事件的场景，至少包括安全风险场景、可解释性风险场景、数据泄露风险场景；
   3) 模型应急规则及决策机制；
   4) 模型应急处置的操作步骤；
   5) 相关文件模版（包括公关文案、法律函件等）。
d) 设置应急复盘机制，对模型生命周期的各个环节进行回溯，确认模型应急事件根本原因和整改方案，同步更新应急预案。
e) 在内部模型监控的基础上，将外部风险指标纳入监控体系，包含客诉、舆情、合作伙伴动态、威胁情报、监管反馈等。
f) 通过红蓝攻防、白帽测试、模型入参干扰等方式进行专项或实战演练。
g) 应急响应平台或工具在第二级的基础上，支持以下功能：
   1) 模型定级；
   2) 模型复盘；
   3) 预案演练；
   4) 应急联动。

参 考 文 献

[1] GB/T 15532—2008 计算机软件测试规范

[2] GB/T 22032—2021 系统与软件工程 系统生存周期过程

[3] GB/T 24363—2009 信息安全技术 信息安全应急响应计划规范

[4] GB/T 25069—2010 信息安全技术 术语

[5] GB/T 32421—2015 软件工程 软件评审与审核

[6] GB/T 32423—2015 系统与软件工程 验证与确认

[7] T/CBA 206—2020 银行业金融机构企业标准体系建设指南

[8] ISO/IEC 15289:2019 Systems and software engineering — Content of life-cycle information items (documentation)

[9] ISO/IEC TR 24030:2021 Information technology — Artificial intelligence (AI) — Use cases

[10] ISO/IEC 27038:2014 Information technology — Security techniques — Specification for digital redaction

[11] ISO 31000:2018 Risk management — Guidelines

[12] ISO 37301:2021 Compliance management systems — Requirements with guidance for use

[13] 《中国银保监会办公厅关于银行业保险业数字化转型的指导意见》（银保监办发〔2022〕2号）